PRELIMINARY DEVELOPMENT OF SOCIAL INDICATOR INDICES

Introduction

It is useful for program planners to have frequency information on all social indicator variables for each of Arizona's fifteen counties. It is also equally useful to see each county's relative frequency on all social indicator variables compared with those of the other fourteen Arizona counties. At some point during the process of examining a county's absolute and relative performance, it is likely that program planners will attempt to move from the simple review of single variables considered on a one-by-one basis to an interpretation that seeks a more global meaning from the range of data and factors presented to them.

Reaching for a more relevant and succinct interpretation is the primary motivation behind the attempt to develop and report on a set of preliminary indices as derived from the larger set of social indicator variables. Such indices should cluster highly related variables into cohesive and similarly interpretable risk groups. They should allow for the computing of an index average – a summarizing of multiple values into a single value – and to provide a simplified and more comprehensive interpretation of the variables.

The derivation of social indicator indices can be accomplished through a systematic and highly visible process, one that is first conceptual and then mathematical. To begin, there must be a grouping of those individual variables that are conceptually linked, which can be readily interpretable as representative of a larger and more inclusive 'idea.' For example, five individual adult arrest variables exist in both the 1997 and 1999 social indicator data sets. It is conceptually possible that these five individual measures all represent related facets of a larger, more inclusive idea of 'adult conflict with the law.' This conceptual linking is frequently based on the literature, for example, following Hawkins and Catalano's model.

Next in the index development process, the variables that are conceptually grouped must provide their own evidence, in the form of bivariate correlation coefficients within the data, demonstrating that positive and significant relationships do exist among the grouped variable sets. Finally, a process of modeling these variable groups is undertaken that will attempt to account for the full set of underlying interrelationships - as expressed in the inter-correlations among variables - and provide a justification for their inclusion into a single index. During this process, individual variables may be retained in the variable set or rejected, on the basis of having or not having consistently positive and important relationships with other variable set members.

Index variables can then be consolidated and summarized. It may be possible to attempt to attach defensible labels as to the increased 'risk' found to be associated with increased values of specific variables, or the increased 'protectiveness' found to be associated with increased values of other variables. The indices themselves will thus provide a more concise picture of a county's health status while highlighting areas of concern, areas of good standing, and areas of change.

A similar process was utilized by the New York State Office of Alcoholism and Substance Abuse Services, which was also funded by SAMHSA to assess their state's prevention needs

through a social indicator study (1997). They developed a set of risk constructs and indices based on locally available variables. The initial step in development of Arizona indices was to examine how Arizona data variables fit into the New York State's Ecological Risk Model. The Arizona 1997 data did not fit the New York State models and we proceeded with further development.

Method

Conceptual process

As stated above, preliminary social indicator indices can be derived through a process that is both systematic and – critically - transparent, regardless of one's level of familiarity with the actual conceptual and mathematical modeling process itself. Individual social indicator variables from the 1997 data set were grouped together in conceptually related clusters on the basis of their interpretability as individual, though representative, measures of larger, more inclusive concepts. Five conceptual indices were thus identified: 1) **Negative Environmental Aspects**, grouping 15 variables; 2) **Educational Status**, grouping 5 variables; 3) **Adult Legal Conflict**, grouping 5 variables; 4) **Youth Legal Conflict**, grouping 8 variables; and 5) **Positive Environmental Aspects**, grouping 5 variables.

The potential variables included in the initial conceptual model were as follows:

1) Negative Environmental Aspects, containing 15 variables:

- Adolescent Pregnancies
- Adolescent Suicides
- Alcohol Sales Outlets
- Alcohol Traffic Fatalities
- Birthrate: Juveniles
- Children Living Away from Home
- Divorce
- Domestic Violence
- Homicides
- Households in Rental
- Mothers Who Used Alcohol During Pregnancy
- Mothers Who Used Tobacco During Pregnancy
- Prisoners
- Single Parent Households
- Tobacco Sales Outlets

2) Educational Status, containing 5 variables:

- Adults without a HS Diploma
- Dropouts Prior to 9th Grade
- Event Dropouts
- Free and Reduced Lunch Program Participants
- Status Dropouts

3) Adult Legal Conflict, containing 5 variables:

- Adult Arrests: Alcohol-Related
- Adult Arrests: Drug-Related
- Adult Arrests: Drunken Driving
- Adult Arrests: Property Crime
- Adult Arrests: Violent Crime

4) Youth Legal Conflict, containing 8 variables:

- Age 10-14 Arrests: Alcohol-Related
- Age 10-14 Arrests: Personal Property
- Age 10-14 Arrests: Vandalism
- Juvenile Arrests: Alcohol-Related
- Juvenile Arrests: Curfew, Vandalism
- Juvenile Arrests: Drug-Related
- Juvenile Arrests: Property
- Juvenile Arrests: Violent Crime

5) Positive Environmental Aspects, containing 5 variables:

- Children in Foster Care
- Employment*
- Net Migration
- New Home Construction
- Population Voting in Elections

*Based on the original variable "Unemployment" (where "Employment" = 100 – "Unemployment")

Mathematical process

Next, the bivariate correlation coefficients of those variables grouped together on a conceptual basis were inspected for demonstrable evidence of the existence of positive and important bivariate relationships; "important" was defined as variable pairs presenting correlation coefficients of at least 0.20 or greater.

Note that at this point in the mathematical process no "inclusion" or "exclusion" decisions (i.e., to retain a variable in a conceptual set or to reject it as being unimportantly related) were made concerning individual variables. Although it was evident for many variables that their relationships to other variables within the set were consistently negative, or were inconsistent in direction, or had correlation coefficients of less than 0.20, decisions were reserved until the actual modeling stage of the process. This 'decision reservation' was made for the simple reason that it is only during actual modeling itself that the multiple bivariate relationships among the conceptual construct can be examined *simultaneously, as a set*, rather than merely on a pair-by-pair, basis.

Lastly, the modeling of the five grouped variable sets or constructs was undertaken. This modeling attempted to provide an accounting of all underlying interrelationships of the full set of

grouped variables. At the same time, the modeling was also an attempt to test the reasonableness of the theory that, in fact, individual variables included in a variable set could be considered as representative and observable measures of a single, more comprehensive, global idea. Successful modeling would thus provide evidence for conceptually and mathematically related variable constructs whose individual members had consistently demonstrated both positive and important inter-relationships. This evidence would be considered a justification for the inclusion of the variable into a single reportable index. During modeling, individual variables were either retained in the variable set or were rejected, all on the basis of their either having or not having consistently positive and process-defined 'important' relationships with other variable set members. For all intents and purposes, this modeling might be termed confirmatory factor analysis.

During this development, it was not the reported variable rates that were of interest, but rather the inter-relationships among the reported variable rates. The correlation matrices of the five grouped variable sets became the basic data of import. The five 1997 grouped variable set correlation matrices are presented below.

Negative Environmental Aspects (15 variables)

	AdlPrg	AdlSui	AlSale	AFatal	AlcPrg	JBirth	ChdAwy	Divorc	DoViol	Homicd	Rental	Prison	SngPar	ToSale	TobPrg
AdlPrg	1.00														
AdlSui	-0.08	1.00													
AlSale	-0.57	-0.44	1.00												
AFatal	0.28	0.00	-0.51	1.00											
AlcPrg	-0.23	0.26	0.07	-0.07	1.00										
JBirth	0.90	0.04	-0.46	0.01	-0.03	1.00									
ChdAwy	0.07	0.39	-0.18	-0.44	0.51	0.32	1.00								
Divorc	0.26	-0.25	-0.09	0.63	0.02	0.13	-0.52	1.00							
DoViol	-0.31	-0.29	0.27	-0.34	0.22	-0.29	-0.16	-0.14	1.00						
Homicd	-0.23	0.59	-0.34	0.11	0.38	-0.19	0.15	-0.27	0.21	1.00					
Rental	-0.21	-0.20	0.06	0.09	-0.53	-0.42	-0.67	-0.09	0.52	0.18	1.00				
Prison	0.08	-0.05	0.33	0.18	-0.05	0.12	-0.28	0.30	-0.22	-0.37	-0.06	1.00			
SngPar	0.24	0.56	-0.24	-0.07	0.21	0.32	0.47	-0.26	-0.28	0.43	-0.41	-0.04	1.00		
ToSale	-0.56	-0.43	0.91	-0.54	0.18	-0.39	-0.03	-0.19	0.27	-0.30	-0.06	0.20	-0.18	1.00	
TobPrg	-0.13	0.11	0.17	0.09	0.47	-0.08	-0.17	0.65	0.05	-0.06	-0.33	0.12	-0.02	0.15	1.00

Educational Status (5 variables)

	NoHSDipl	9Dropout	EDropout	FreeRedc	SDropout
NoHSDipl	1.00				
9Dropout	0.42	1.00			
EDropout	0.34	0.62	1.00		
FreeRedc	0.44	0.50	0.52	1.00	
SDropout	0.42	0.70	0.66	0.42	1.00

Adult Legal Conflict (5 variables)

	Alcohol	Property	Violent	Drug	Drunken
Alcohol	1.00				
Property	0.76	1.00			
Violent	0.64	0.62	1.00		
Drug	0.70	0.53	0.57	1.00	
Drunken	0.93	0.64	0.71	0.73	1.00

Youth Legal Conflict (8 variables)

	1014Alchl	JuvAlchl	Curfew	Property	Violent	Drug	1014Prop	Vandal
1014Alchl	1.00							
JuvAlchl	0.98	1.00						
Curfew	0.02	0.02	1.00					
Property	0.23	0.18	0.68	1.00				
Violent	0.04	-0.01	0.47	0.61	1.00			
Drug	0.21	0.24	0.37	0.43	0.77	1.00		
1014Prop	0.18	0.15	0.70	0.98	0.51	0.37	1.00	
Vandal	0.22	0.19	0.72	0.84	0.47	0.44	0.87	1.00

Positive Environmental Aspects (5 variables)

	FostCare	Migratn	NewHome	PopVote	Employmt
FostCare	1.00				
Migratn	0.32	1.00			
NewHome	0.49	0.44	1.00		
PopVote	0.56	0.33	0.33	1.00	
Employmt	0.50	0.19	0.42	0.47	1.00

Modeling Criteria/Assumptions

Several criteria were established prior to the beginning of modeling. The retention of individual variables in a grouped variable set and the retention of the overall variable set model itself. Individual variables with either negative factor loadings or loadings of less than 0.20 were to be excluded from a variable set, with modeling to be resumed using the reduced variable set. The p-value level was set at p >/= .05 for model retention purposes. The second model retention criterion – that the model's root mean square error of approximation (RMSEA) must reach a level no higher than 0.10 – was also employed. This second model retention criterion provided a criterion that was sample-size independent. Finally, variable set correlation matrices were assumed to be based on 200 observations: Larger numbers of observations are likely to make modeling oversensitive to minor deviations between estimated and observed correlation matrices, while smaller numbers of observations are likely to inflate the statistical significance of estimated model parameters.

Model Confirmation/Validation using the 1999 Social Indicator Data Set

After the above modeling of the grouped variable constructs was completed using the 1997 data set, the five obtained variable group model structures were then tested using correlation matrices derived from the 1999 social indicator data set. In line with modeling criteria stated above,

individual variables from the 1999 data set with either negative factor loadings or loadings of less than 0.20 were excluded, with modeling then resumed using the reduced variable set. The corresponding 1997 model was then also adjusted so that any variable showing relational instability was removed.

Results

The following five variable set models were obtained through the conceptual and mathematical modeling processes described above using the 1997 data set; these models were then retained, after the 1999 modeling, with only one minor adjustment with the final exclusion of the "Homicides" variable from the **Negative Environmental Aspects** model.

1) Negative Environmental Aspects (4 retained variables common to 1997 and 1999; 6 retained variables for 1999 data set)

- Adolescent Suicides
- Children Living Away from Home
- Food Stamps*
- Mothers Who Used Alcohol During Pregnancy
- Single Parent Households
- TANF*

* Available only in the 1999 data set

1997 Data Set: Correlation Matrix Analyzed

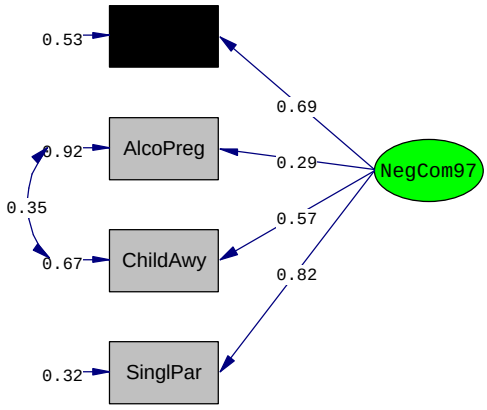
	AdolSuic	AlcoPreg	ChildAwy	SinglPar
AdolSuic	1.00			
AlcoPreg	0.26	1.00		
ChildAwy	0.39	0.51	1.00	
SinglPar	0.56	0.21	0.47	1.00

1999 Data Set: Correlation Matrix Analyzed

	AdolSuic	AlcoPreg	ChildAwy	SinglPar
AdolSuic	1.00			
AlcoPreg	0.67	1.00		
ChildAwy	0.61	0.71	1.00	
SinglPar	0.31	0.24	0.47	1.00

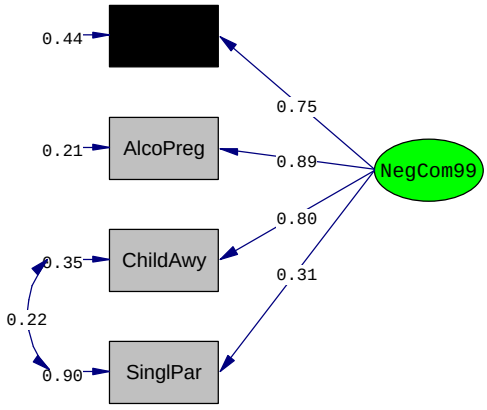
1999 Data Set: Correlation Matrix Analyzed (including the two new variables, "Food Stamps" and "TANF" available only in the 1999 data set)

	AdolSuic	AlcoPreg	ChildAwy	SinglPar	FoodStmp	TANF
AdolSuic	1.00					
AlcoPreg	0.67	1.00				
ChildAwy	0.61	0.71	1.00			
SinglPar	0.31	0.24	0.47	1.00		
FoodStmp	0.53	0.69	0.84	0.29	1.00	
TANF	0.56	0.77	0.85	0.36	0.98	1.00



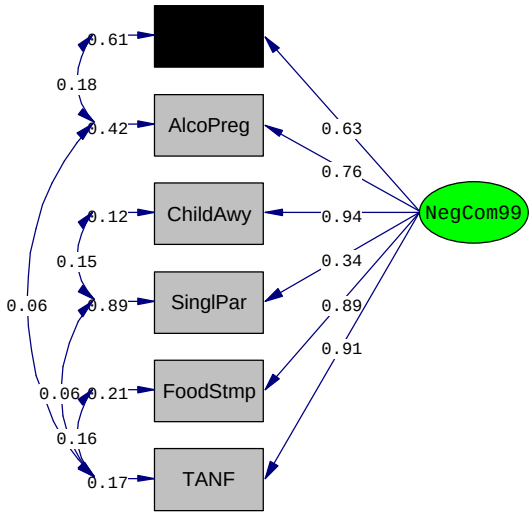
Chi-Square=2.87, df=1, P-value=0.08999, RMSEA=0.097

Negative Environmental Aspects
1997 Data Set



Chi-Square=4.17, df=1, P-value=0.04122, RMSEA=0.126

Negative Environmental Aspects
1999 Data Set



Chi-Square=6.56, df=4, P-value=0.16095, RMSEA=0.057

Negative Environmental Aspects
1999 Data Set (including two new variables, "Food Stamps" and "TANF", available only in the 1999 data set)

2) Educational Status (5 retained variables)

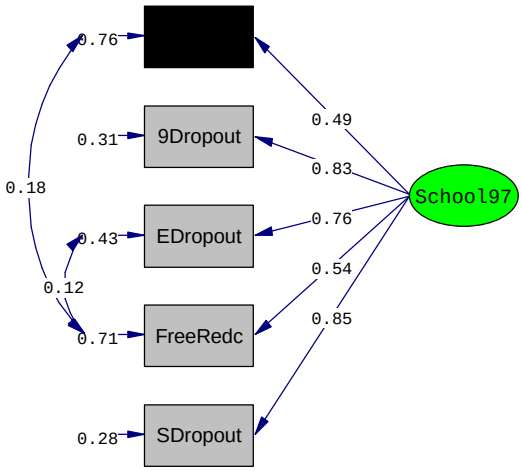
- Adults without a HS Diploma
- Dropouts Prior to 9th Grade
- Event Dropouts
- Free and Reduced Lunch Program Participants
- Status Dropouts

1997 Data Set: Correlation Matrix Analyzed

	NoHSDipl	9Dropout	EDropout	FreeRedc	SDropout
NoHSDipl	1.00				
9Dropout	0.42	1.00			
EDropout	0.34	0.62	1.00		
FreeRedc	0.44	0.50	0.52	1.00	
SDropout	0.42	0.70	0.66	0.42	1.00

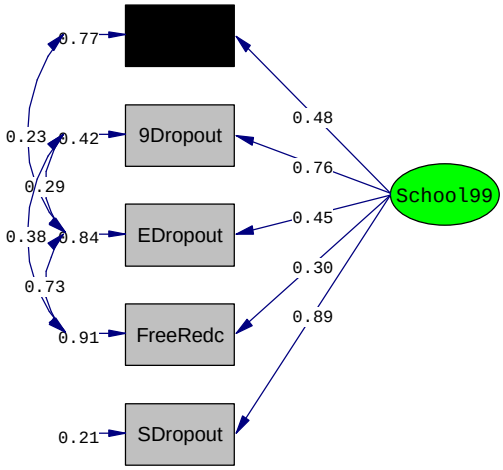
1999 Data Set: Correlation Matrix Analyzed

	NoHSDipl	9Dropout	EDropout	FreeRedc	SDropout
NoHSDipl	1.00				
9Dropout	0.34	1.00			
EDropout	0.37	0.63	1.00		
FreeRedc	0.05	0.61	0.83	1.00	
SDropout	0.42	0.68	0.41	0.28	1.00



Chi-Square=6.23, df=3, P-value=0.10086, RMSEA=0.074

Educational Status
1997 Data Set



Chi-Square=2.92, df=1, P-value=0.08758, RMSEA=0.098

Educational Status
1999 Data Set

3) Adult Legal Conflict (5 retained variables)

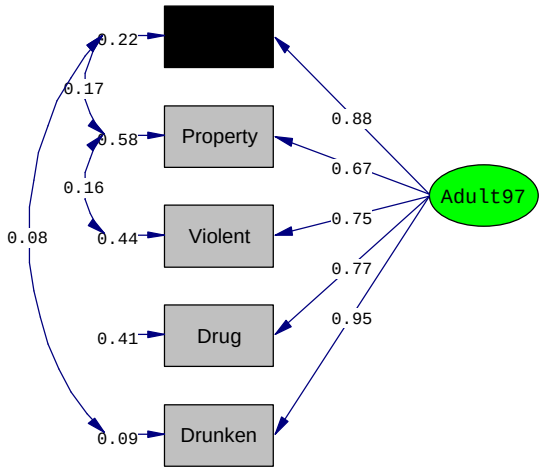
- Adult Arrests: Alcohol-Related
- Adult Arrests: Drug-Related
- Adult Arrests: Drunken Driving
- Adult Arrests: Property Crime
- Adult Arrests: Violent Crime

1997 Data Set: Correlation Matrix Analyzed

	Alcohol	Property	Violent	Drug	Drunken
Alcohol	1.00				
Property	0.76	1.00			
Violent	0.64	0.62	1.00		
Drug	0.70	0.53	0.57	1.00	
Drunken	0.93	0.64	0.71	0.73	1.00

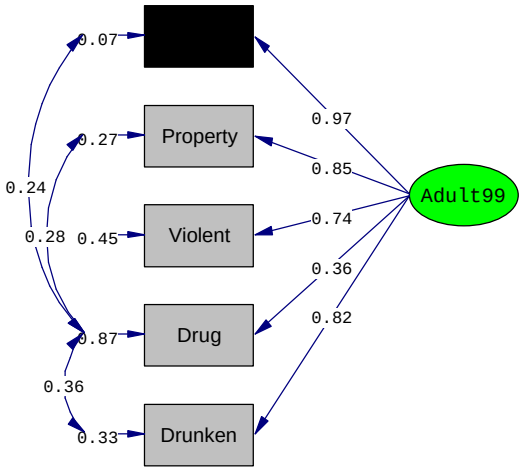
1999 Data Set: Correlation Matrix Analyzed

	Alcohol	Property	Violent	Drug	Drunken
Alcohol	1.00				
Property	0.82	1.00			
Violent	0.71	0.66	1.00		
Drug	0.59	0.57	0.27	1.00	
Drunken	0.79	0.68	0.59	0.65	1.00



Chi-Square=3.28, df=2, P-value=0.19440, RMSEA=0.057

Adult Legal Conflict
1997 Data Set



Chi-Square=2.28, df=2, P-value=0.32015, RMSEA=0.026

Adult Legal Conflict
1999 Data Set

4) Youth Legal Conflict (5 retained variables)

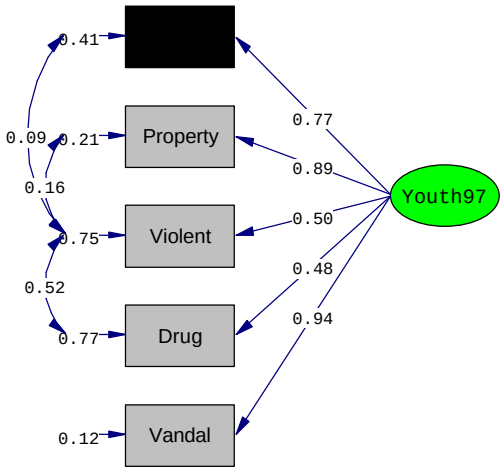
- Age 10-14 Arrests: Vandalism
- Juvenile Arrests: Curfew, Vandalism
- Juvenile Arrests: Drug-Related
- Juvenile Arrests: Property
- Juvenile Arrests: Violent Crime

1997 Data Set: Correlation Matrix Analyzed

	Curfew	Property	Violent	Drug	Vandal
Curfew	1.00				
Property	0.68	1.00			
Violent	0.47	0.61	1.00		
Drug	0.37	0.43	0.77	1.00	
Vandal	0.72	0.84	0.47	0.44	1.00

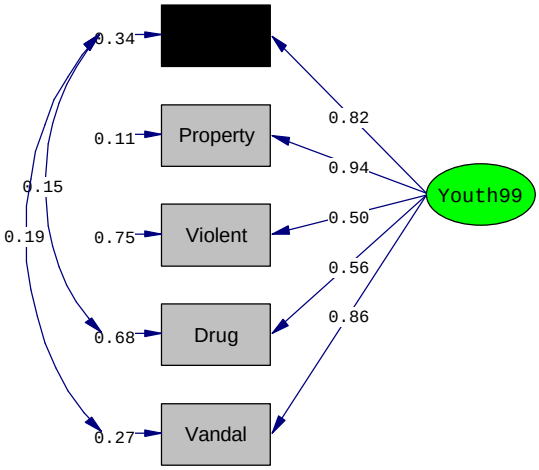
1999 Data Set: Correlation Matrix Analyzed

	Curfew	Property	Violent	Drug	Vandal
Curfew	1.00				
Property	0.77	1.00			
Violent	0.44	0.47	1.00		
Drug	0.61	0.53	0.34	1.00	
Vandal	0.89	0.81	0.42	0.48	1.00



Chi-Square=0.12, df=2, P-value=0.94320, RMSEA=0.000

Youth Legal Conflict
1997 Data Set



Chi-Square=3.31, df=3, P-value=0.34575, RMSEA=0.023

Youth Legal Conflict
1999 Data Set

5) Positive Environmental Aspects (5 retained variables)

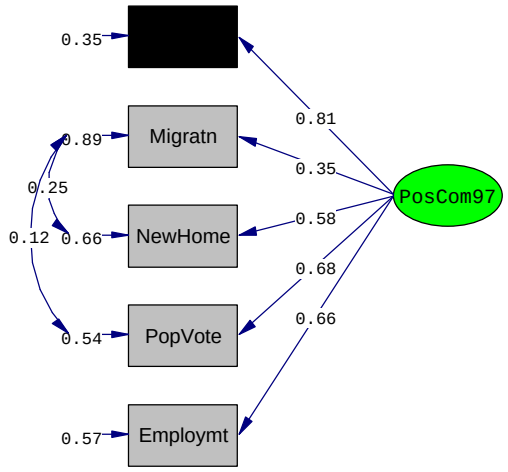
- Employment
- Foster Care
- Net Migration
- New Home Construction
- Population Voting in Elections

1997 Data Set: Correlation Matrix Analyzed

	FostCare	Migratn	NewHome	PopVote	Employmt
FostCare	1.00				
Migratn	0.32	1.00			
NewHome	0.49	0.44	1.00		
PopVote	0.56	0.33	0.33	1.00	
Employmt	0.50	0.19	0.42	0.47	1.00

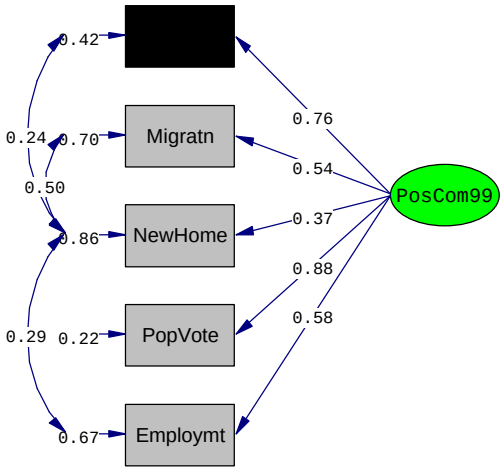
1999 Data Set: Correlation Matrix Analyzed

	FostCare	Migratn	NewHome	PopVote	Employmt
FostCare	1.00				
Migratn	0.44	1.00			
NewHome	0.54	0.71	1.00		
PopVote	0.67	0.47	0.33	1.00	
Employmt	0.43	0.28	0.48	0.52	1.00



Chi-Square=6.54, df=3, P-value=0.08823, RMSEA=0.077

Positive Environmental Aspects
1997 Data Set



Chi-Square=1.37, df=2, P-value=0.50346, RMSEA=0.000

Positive Environmental Aspects
1999 Data Set

Recommendations

While the social indicator indices presented in this report have been labeled, they must be considered as only preliminary in nature. Numerous data set limitations exist as to whether:

- all pertinent variables have indeed been collected;
- collected variables have been measured with sufficient precision to enable sound modeling to occur;
- collected variables have been measured on a requisite level (be it county, city/town, or zip code area) to allow for sufficient modeling precision.

The following recommendations are made regarding future index modeling:

- Use data representing a longer data collection period; for example, a five-year collection period rather than single year collection periods.
- Conduct the remodeling using updated 2000 census variables.
- Consider excluding either “Food Stamps” or “TANF” from future iterations of the **Negative Environmental Aspects** model, due to high variable inter-correlation (correlation = 0.98 in the 1999 data set), although both variables were of sufficient loading in the model tested.

- Focus on the collection of variables that are readily obtainable, that can be obtained with a determinable level of precision, and that have identical (or highly similar) variable counterparts in other states.

- Attempt to develop additional social indicator variables that capture more of the information available from **Negative Environmental Aspects** variables rejected during the present modeling.

- Experiment with alternative index constructs that might provide equally useful variable set interpretations.

Guide to Social Indicator Indices Interpretation

Preliminary social indicator indices consist of conceptually and mathematically related variables that have been consolidated and summarized into single index values. These indices thus simplify variable value interpretation as well as enable a clearer and more supportable interpretation of the risk or protective characteristics associated with increased values of identified index variables. Improved variable interpretation will enhance the ability of reviewers to more soundly determine a county's particular health status, in terms of county concerns, county successes, and areas in which a county appears to be undergoing change.

Using the actual social indicator indices values based on the 1999 data set and herein reported for Apache County, the following five index interpretations have been provided as examples of standard SIS indices interpretation.

1) Apache County average z-score for the **Negative Environmental Aspects** variables: **1.97**
   Interpretation: Apache County has, on average, a risk value for **Negative Environmental Aspects** that is 1.97 standard deviations above the all-counties average risk value (i.e., it has a greater risk here than do other counties).

2) Apache County average z-score for the **Educational Status** variables: **0.77**
   Interpretation: Apache County has, on average, a risk value for **Educational Status** that is 0.77 standard deviations above the all-counties average risk value (i.e., it has a greater risk here than do other counties).

3) Apache County average z-score for the **Adult Legal Conflict** variables: **-1.40**
   Interpretation: Apache County has, on average, a risk value for **Adult Legal Conflict** that is 1.40 standard deviations below the all-counties average risk value (i.e., it has a lesser risk here than do other counties).

4) Apache County average z-score for the **Youth Legal Conflict** variables: **-1.45**
   Interpretation: Apache County has, on average, a risk value for **Youth Legal Conflict** that is 1.45 standard deviations below the all-counties average risk value (i.e., it has a lesser risk here than do other counties).

5) Apache County average z-score for the **Positive Environmental Aspects** variables: **-1.27**
   Interpretation: Apache County has, on average, a protective value for **Positive Environmental Aspects** that is 1.27 standard deviations below the all-counties average protective value (i.e., it has a lesser protective value here than do other counties).